

Constitutional Review via Facebook's Oversight Board

Lorenzo Gradoni

2021-02-10T12:23:06

Facebook's Oversight Board (OB) "[is unlike anything previously created by a company](#)", and "[a leap into the unknown](#)". A [trust](#) and a [limited liability company](#), both incorporated in Delaware, make its existence possible. But the OB is neither that trust nor that company, nor the two combined. It is, rather, an arresting example of [societal constitutionalism](#), with a serious claim to global lawmaker status.

The OB is, in essence, an adjudicator – a "Supreme Court of Facebook", as Professor Noah Feldman dubbed it in a seminal [memo](#) that reached Menlo Park in early 2018. "Imagine some sort of structure, almost like a Supreme Court...", [echoed](#) Zuckerberg, a few months later. The OB is formally independent of Facebook (no "F" in the acronym, then) and its decisions bind its creator. Currently, the OB's jurisdiction is limited to disputed takedowns of single posts or comments. However, no such limit exists where Facebook itself makes a referral, which is how the [Trump deplatforming affair](#) fell into the OB's lap — a case of incalculable significance that the OB will likely settle applying (hang on to your hats) public international law.

For all its disorienting novelty, however, the OB's recent [debut](#) thrust us back to a familiar script, that of the emergence of constitutional review, the power reclaimed by the US Supreme Court in [Marbury v Madison](#). The resemblance not only passed unnoticed; commentators emphatically deny it. For scholars at Lawfare, which hosts a "[blog in the blog](#)" entirely devoted to Facebook's "grand experiment", the OB's inaugural jurisprudence is "not exactly comparable to the Supreme Court's first-ever ruling in *West v. Barnes*, let alone *Marbury v. Madison*". Writing for Verfassungsblog, Oreste Pollicino and Giovanni De Gregorio [concurred](#). I pick up on this casually-made and *strangely recurrent* historical parallel and argue that at closer inspection, the OB's first set of decisions are, indeed, the *Marbury v Madison* of platform governance.

In its all-time classic, the US Supreme Court, without express constitutional warrant, affirmed its power to void unconstitutional laws. The OB already reached a comparable level of ambition, using international human rights as a lever. In the face of Facebook's enduring qualms about human rights, the OB single-mindedly harnessed their power, changing Facebook's normative system almost beyond recognition. On the day of the OB's still underrated debut, human rights inexorably emerged as the main touchstone of Facebook's performance *as legislator*, over and above its role of enforcer of online speech rules.

Lex Facebook and the Spectre of Human Rights

In [Lex Facebook](#), prescription and persuasion intermingle: dry precepts and didascallic remarks sit side by side, preambles give way to introductions, and learn-more-about hyperlinks replace cross-references. Beneath such coating of customer-service language, however, it is easy to recognize a hierarchically-structured [normative system](#) with supreme principles at its apex, called Facebook's Values (among which "Voice" is paramount), an intermediate level of repressive norms — the Community Standards — and a lower stratum teeming with detailed directives. These are the Internal Implementation Standards, instructions crafted to enable high-speed micro-deliberation by an army of human moderators, and the code informing AI-driven content skimming. Values and Community Standards are public, Internal Implementation Standards and the AI protocols are not. Moderators have occasionally [leaked](#) the rulebook, though, and, as we shall see, the OB may let bits of it come to the surface.

Human rights are conspicuously absent from this body of rules, and yet they inhabit it like ghosts. Although *Lex Facebook* never references them, the [introduction](#) to the Community Standards explains that they affect the decision-making process through "the advice of experts". Furthermore, where Facebook determines that an otherwise impermissible piece of content must stay up in the public interest, it does so only "after weighing the public interest value against the risk of harm", "look[ing] to international human rights standards" for guidance. Facebook's Values are, more than anything else, a set of idiosyncratically-phrased principles that ground a sort of boardroom human rights adjudication.

Facebook's approach of not referencing human rights (and external normative sources in general) didn't change, as it set about to devolve part of its adjudicative jurisdiction. Disclosed in January 2019, the OB's [Draft Charter](#) — actually a discursive set of suggestions — failed to mention human rights where it tangentially dealt with relevant "values". Released on 17 September 2019, the [Charter](#) — "[the foundational governing document for the Oversight Board](#)" — features scant references to them. And yet, the issue of human rights had persistently come up during the public consultation process orchestrated by Facebook in the lead-up to OB's inception. The [consultation report](#) records that "at the vast majority of workshops and roundtables", participants "proposed that the Board incorporate international human rights norms and standards into its core decision-making functions". Facebook didn't budge.

That Facebook chose to keep human rights at arm's length shouldn't come as a surprise. During the last half-century, transnational corporations have fiercely opposed the idea that international human rights norms formally bind them. Facebook, despite its penchant for lofty declarations, is no exception. However, as a member of the [Global Network Initiative](#), Facebook has arguably committed to acting consistently with the [UN Guiding Principles on Business and Human Rights](#) (UNGPs). The UNGPs play a crucial role in this story. They are the [master move](#) by which [John Ruggie](#), in the elusive Business & Human Rights chessboard, tied corporate governance logics to public international legal discourse,

particularly by reframing compliance with human rights as a means of business risk management.

Indeed, the OB's Charter sticks to a risk-avoidance approach – one that is seemingly confined to breaches of freedom of expression. It stipulates that “[w]hen reviewing decisions, the board will pay particular attention to the impact of removing content in light of human rights norms protecting free expression” (Art. 2.2). The OB Charter conceives human rights as a mere complement to *Lex Facebook*, not a superior source of law. The OB's core function, as defined by the Charter, accordingly consists in “review[ing] content enforcement decisions and determine whether they were consistent with Facebook's content policies and values”. In so doing, the OB will have to “*interpret* Facebook's Community Standards [...] in light of Facebook's articulated values” (Art. 1.4.2, Facebook's emphasis). Under the Charter, the OB must then pass judgment *based on Lex Facebook*. It may not criticize *Lex Facebook*, except where it offers policy advice that the company is ultimately free to ignore (Art. 1.4 and 3.7.3).

The [OB's Bylaws](#) feature an admirably concise restatement of *Lex Facebook's* self-contained character, and the OB's duty to keep review focused on individual decisions about content: “The board will review and decide on content in accordance with Facebook's content policies and values” (Art. 1.3). On top of this, Facebook took special care to prevent the OB from relying on anything resembling the implied powers doctrine: “The board will have no authority or powers beyond those expressly defined by this charter” (Art. 1.4). The drafters of the US Constitution were less alert to the risk of judicial activism. Thus, the OB had to overcome a higher interpretative obstacle to break into the business of constitutional review. Well, it did.

Human Rights Creeping In (and Taking Over)

Shortly after its establishment, the OB espoused an expansive conception of human rights' significance for the discharge of its mission, so unlike Facebook's. The OB [introduced itself](#) to the world as “an organization with a global scope and focus on free expression and human rights”. While announcing its readiness to receive appeals, the OB [disclosed](#) plans to check Facebook's takedowns against “international human rights norms and standards”. Soon after, it anchored human rights in its first act of self-regulation, the [Rulebook for Case Review and Policy Guidance](#), wherein it also referenced the UNGPs.

On 21 January 2021, explaining its eagerness to accept Facebook's referral of the Trump deplatforming affair, the OB [claimed](#) that it “was created to provide a critical, independent check on Facebook's approach to the most challenging content issues, which have enormous implications for global human rights and free expression”. Now, “Facebook's approach” is a concept more sweeping than that of individual moderation decisions and certainly embraces what Facebook does in its legislative capacity. An oversight? Not really.

The moment of truth arrived as the OB discussed what appeared to be the most straightforward case in the first batch (Case Decision [2020-004-IG-UA](#)). Facebook

had removed pictures of uncovered female nipples posted on Instagram as part of a campaign to raise awareness of breast cancer symptoms. Subsequently, Facebook acknowledged the machine-made mistake and reinstated the post, inviting the OB to drop the case. The OB declined, arguing that doing so deprives affected users of their right to a remedy under Art. 2 of the [International Covenant on Civil and Political Rights](#) (ICCPR), in the form of “a fully reasoned decision”. The OB then went on to make of this case, regarded by Facebook as moot, nothing less than its *Marbury v Madison*.

Facebook insisted that it was “not relevant to the Board’s consideration of the case whether the content was removed through an automated process, or whether there was an internal review to a human moderator”. To this it added, with revealing rudeness: “Facebook would like the Board to focus on the outcome of enforcement, and not the method”. Note that “the method” is *Lex Facebook*. By refusing to oblige, therefore, the OB affirmed its power to assess whether Facebook’s normative output complies with human rights.

The structure of the OB’s reasoning is identical across decisions. First, the OB checks Facebook’s behavior against the Community Standard and the Values. It then switches to international human rights law. The switch, remarkably, is enabled by the circumstance that Facebook endorsed the UNGPs, “a voluntary framework for businesses’ human rights responsibilities” which conjures up a panoply of human right instruments, binding and non-binding (see, e.g., Case Decision [2020-005-FB-AU](#)). As the OB breathtakingly remarked, Principle 31 of the UNGPs “specify that non-judicial grievance mechanisms (such as the Oversight Board) should deliver outcomes that accord with internationally recognized human rights” (Case Decision [2020-006-FB-FBR](#)). Human rights analysis forms both the decisions’ core and their most extensive section. In comparison, analysis under the Values is ultracompact (sometimes consisting of just a few lines) and made redundant by the human rights section, which is where the real balancing happens. The OB has *de facto* neutralized the Values and enthroned human rights.

Constitutional Review in Full Swing

With its first decisions, the OB engaged in a sustained critique of Facebook-as-legislator. Most of this critique brings into play the tripartite test that human rights jurisprudence draws from [Art. 19.3 ICCPR](#), under which only restrictions of speech prescribed by law for a legitimate purpose, and necessary to achieve that purpose, are lawful. Particularly under the legality test, the OB spotted shortcomings at all levels of *Lex Facebook*. It didn’t spare the Community Standards – recall that under the Charter it should just apply them –, finding some of them “inappropriately vague”, even inconsistent, or a confusing “patchwork of rules and policies” scattered all over the website (Case Decisions [2020-005-FB-UA](#) and [2020-006-FB-FBR](#)).

In one case, the OB unearthed and voided an Internal Implementation Standard – these are not public knowledge – which mandated the removal of quotes attributed to “dangerous individuals”, unless the user clarified that quoting did not mean praise or support. For the OB, such shortcuts are inadmissible: requiring moderators to

disregard other “contextual cues” entails “an unnecessary and disproportionate restriction on expression” (Case Decision [2020-005-FB-UA](#)).

In the above-mentioned [breast cancer case](#), the OB knocked down pieces of AI code, noting in passing that the latter form an integral part of *Lex Facebook* (“the content policies are essentially embedded into code and be considered inseparable from it and self-enforcing”). In the OB’s view, “automated content moderation without necessary safeguards is not a proportionate way for Facebook to address violating forms of adult nudity”. Worse still, the code is bound to discriminate against women: “Given that Facebook’s rules treat male and female nipples differently, the reliance on inaccurate automation to enforce those rules will likely have a disproportionate impact on women”.

It is easy to see, in retrospect, that the activity of adjudicating upon alleged breaches of *Lex Facebook* would have led, sooner or later, to findings of structural violations — violations stemming from flaws in the law and not just from flawed law-enforcement. The OB detected a few such violations at its debut. It may have looked the other way. Instead, it decided to weed them out armed with human rights law. [Facebook](#) should relish Justice Marshall’s big smile on its creature’s face. Wasn’t it a Supreme Court, after all?

